Corpus and Computational Tools for Generative Metrics

Stephanie Sin-yun Shih
Stanford University

Symposium: A comparison of models for meter: Corpora and other sources of evidence for metrical theory and method

Linguistic Society of America Annual Meeting
Baltimore, Maryland
January 8, 2009

The Metrical Corpus

- Shakespeare = 118, 406 lines

- Construction of a metrically-parsed corpus of poetic texts

  → Quantitative data for evaluation of metrical hypotheses
- Building and designing an automatic metrical annotator/parser
- Preliminary results, problems, future questions

Goal:
The Metrical Corpus

- Generative/Parametric Theory (Jespersen 1933, Hanson & Kiparsky 1996, a.o.):
  - Metrical Template
    (w s)(w s)(w s)(w s)(w s) [=Iambic pentameter]
  - Correspondence constraints:
    - Matches input to metrical template
    - Polysyllabic main stressed syllables must be in metrical 's'

- Music to hear, why hear'st thou music sadly?
  (w s) (w s) (w s) (w s)(w s) <>

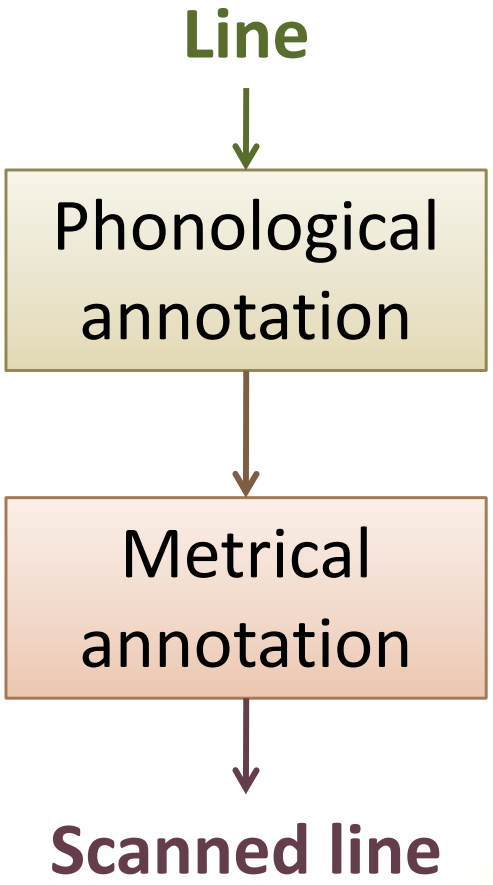
→ Metrical corpus = scanned lines of poetic text
(akin to a syntactically parsed corpus, e.g., Treebank)

Why?
The Metrical Corpus

- Gold-standard corpus
- Standardization within the theory
- Handling and analysis of quantitative data
- Inter-coder reliability
- (Convenience)

Getting there:
Automatic Annotation

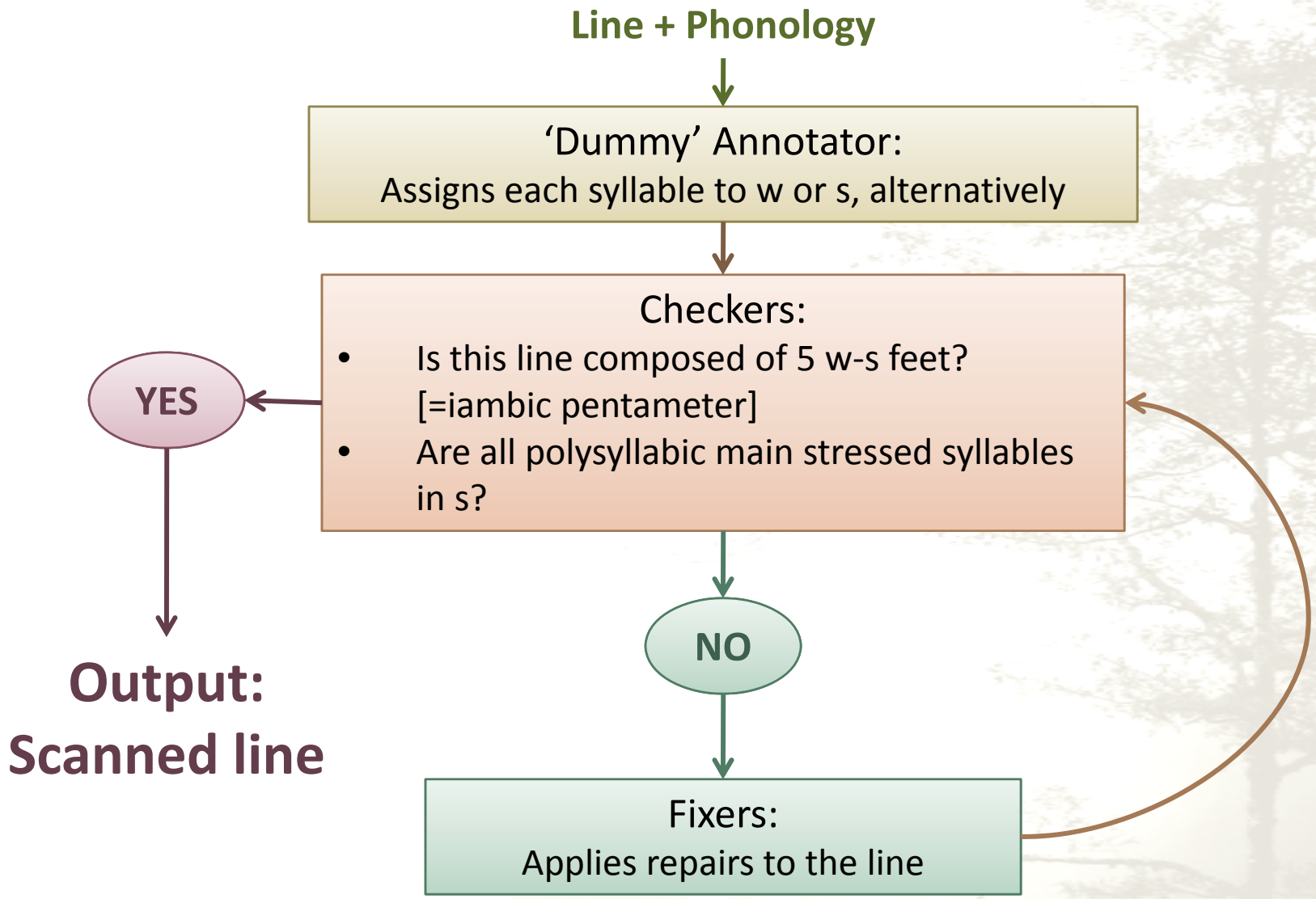
- Python code
- Two-step annotation process:



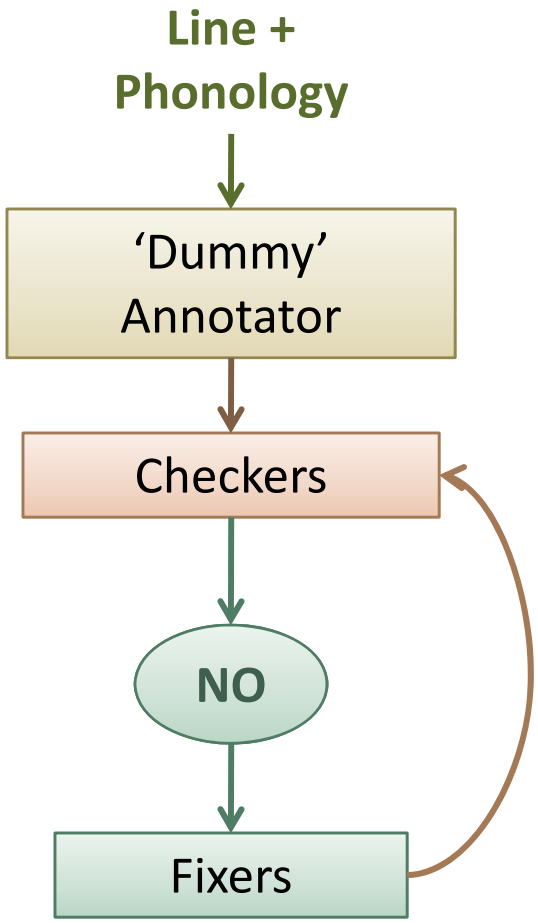
Stress, weight, segments:

1. Automatic annotation from Carnegie Mellon University online dictionary and Speriosu 2007
2. ~700 hand-annotated words for stress and weight
3. Override dictionary (pronouns, e.g., ‘me’, ‘ye’)

Automatic Metrical Annotation

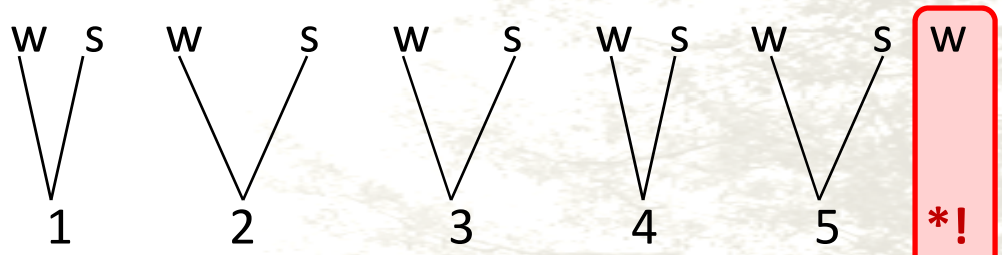


Example:
Automatic Metrical Annotator



P U P P U P U P U P U

And in fresh numbers number all your graces



(fail: more than 5 w-s feet in line)

w s w s w s w s w s <>

(extrametricality fixer: extra line-final syllable allowed if it is U.)

Automatic Metrical Annotator:
Output

	line_num	line	meter	stress	weight	comment
2	1:01	From fairest creatures we desire increase,	w sw sw s ws ws	P PU PU P UP UP	H LH HH H LH HH	CLEAR
3	1:02	That thereby beauty's rose might never die,	w sw sw s w sw s	P PP PU P P PU P	H HH HH H H LH H	CLEAR
4	1:03	But as the riper should by time decease,	w s w sw s w s ws	P P U PU P P P UP	H H L HH H H H LH	CLEAR
5	1:04	His tender heir might bear his memory:	w sw s w s w sws	P PU P P P P PUU	H HH H H H H LHH	CLEAR
6	1:05	But thou, contracted to thine own bright eyes,	w s Wsw s w s w s	P P PUU P P P P P	H H HHH H H H H H	| Internal Inversion
7	1:06	Feed'st thy light's flame with self-substantial fuel,	w s w s w swsw s<>	P P P P P SUPU PU	H H H H H HHHH HH	| Extrametrical ending
8	1:07	Making a famine where abundance lies,	Ws w sw s wsw s	PU U PU P UPU P	HH L LH H LHH H	| Initial Inversion
9	1:08	Thyself thy foe, to thy sweet self too cruel.	ws w s w s w s w s<>	UP P P P P P P P PU	HH H H H H H H H HH	| Extrametrical ending
10	1:09	Thou that art now the world's fresh ornament,	w s w s w s w sws	P P P P U P P PUU	H H H H L H H HLH	CLEAR
11	1:10	And only herald to the gaudy spring,	w sw sw s w sw s	P PU PU P U PU P	H HH LH H L HH H	CLEAR
12	1:11	Within thine own bud buriest thy content,	ws w s w sws w s<>	UP P P P PUU P PU	LH H H H LHH H HH	| Extrametrical ending
13	1:12	And, tender chorl, mak'st waste in niggarding:	w sw s w s w sws	P PU P P P U PUU	H HH H H H H LHH	CLEAR
14	1:13	Pity the world, or else this glutton be,	Ws w s w s w sw s	PU U P P P P PU P	LH L H H H H HH H	| Initial Inversion
15	1:14	To eat the world's due, by the grave and thee.	w s w s w s w s w s	P P U P P P U P P P	H H L H H H L H H H	CLEAR

96	The eyes ('fore duteous) now converted are	w s w sws w sws w	U P P PUU P UPU P	| Main stress in strong: FAIL | Foottype or num problem: FAIL
97	From his low tract and look another way:	w s w s w s wsw s	P P P P P P UPU P	CLEAR
98	So thou, thyself outgoing in thy noon,	w s ws wsw s w s	P P UP PSU U P P	| Main stress in strong: FAIL
99	Unlook'd on diest unless thou get a son.	ws w sw sw s w s w	UP P PU UP P P U P	| Main stress in strong: FAIL | Foottype or num problem: FAIL

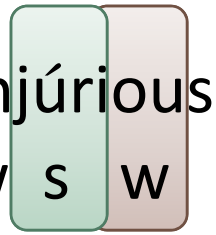
Automatic Metrical Annotator:
Fixer Module

1. Line-final extrametricality
2. Line-internal extrametricality
3. Line-initial inversion
4. Line-internal inversion
5. Resolution
6. Elision
7. Sesquipsyllabic contraction
8. Trisyllabic substitution
9. Squeezing of unstressed function words

Automatic Metrical Annotator:
Fixer Module Problems

Overlapping repair strategies

- E.g.,
  - *Resolution*: V C V allowed in a single metrical S
    - e.g., prodigal
  - *Elision*: adjacent vowels may share a metrical pos
    - e.g., being



Injúrious dístance should not stop my way,

w s w s w s w s w s

→ How much information is necessary to identify and distinguish iambic pentameter?

Automatic Metrical Annotator:
Fixer Module Problems

Interaction of repairs:

- What order do the repairs occur in? How does the annotator apply repairs to achieve an optimal line?
- Potential solution = Optimality-theoretic model (?)
- GEN: generating all possible repaired outputs of a problematic line

Automatic Metrical Annotator:
Baseline and Improvements

(Total # of lines = 2155)	# annotated lines	# total annotated lines	% of total lines
'Dummy' annotation	1544	1544	71.65
+ Line-final extrametricality	+ 162	1706	79.16
+ Line-initial inversion	+ 102	1808	83.9

- Random sample of 'dummy' annotated lines hand-checked for accuracy:

	# of checked lines	# of errors	Error %
'Dummy' annotation	402	9	2.21

- Extrametricality, inversion hand-checked for accuracy:

	Total # of lines	# of errors	Error %
Line-final extrametricality	167	4	2.4
Line-initial inversion	102	1	< 1

Metrical Corpus | Preliminary Results:
Syllable Weight in Inversion

- Initial inversion:
- Músic to hear, why hear'st thou music sadly?

  (w s) (w s) (w s) (w s)(w s) <>

  *(s w)(w …
- Kiparsky 2005: inverted iambs ≠ trochees
  - Inverted iambs prefer L σ; Trochees prefer H σ

Metrical Corpus | Preliminary Results:
Syllable Weight in Inversion

	Line-initial Inversion	All initially-stressed words
Heavy initial syllable	73 (=71.57%)	928 (=65.26%)
Light initial syllable	29 (=28.43%)	494 (=34.74%)
	102	**1422**

- Shakespeare's sonnets: more heavy syllables than light ones in polysyllabic initial inversion, counter to Kiparsky (2005)'s Finnish findings
- Inverted iambs ≠ trochees ?
- Hanson 2009: Inversions "[serve] to signal strong the beginning of a new line" (281) in English and Romance meter.

Conclusion

- Goal: Metrically-parsed corpus with scanned lines of poetic text
- Automatic metrical annotator: given phonology and verse, metrically parses the input

- Eventually: annotator applicable cross-linguistically to (metrical) verse forms

Conclusion:
Automatic Metrical Annotator

- Automatic metrical annotator
  - ~84% of lines annotated
  - Baseline from 'dummy' annotation ~72%
  - Low error rate ~2%

- Implementation of metrical 'repairs,' most likely using an Optimality-based approach
- Further annotations:
  - Phonological segments
  - Syntactic information
  - Phrasal stress
  - Optional/alternative pronunciations

Conclusion:
Metrical Corpus

- Preliminary quantitative data with line-initial inversions shows differences in weight distribution in English and Finnish.

- Ultimately:
  - Distribution of irregularities in the meter
  - What and how much does an automatic annotator need to know to produce ‘correct’ or ‘optimal’ scansions of poetry?

    => How much do we know to intuitively perceive meter?

Thank you!

- Thank you to…
  - …San Duanmu and Nigel Fabb for organizing this symposium.
  - …Arto Anttila, Paul Kiparsky, and the Stanford University Phonetics and Phonology Workshop for their input, questions, comments, and posers.
  - …Mike Speriosu, Richard Conway, and Chris Potts for invaluable programming aid and advice.
  - …Kristin Hanson for assigning tedious metrics homework that inspired this project.

Select References

Anttila, Arto and Mike Speriosu. 2007. "Notes on English stress and weight annotation." <http://stanford .edu/~anttila/research/stress-weight-notes-jan-2007.pdf>

Fabb, Nigel and Morris Halle. 2008. *Meter in Poetry: A New Theory.* Cambridge: Cambridge University Press.

Hanson, Kristin. 2009. "Metrical alignment." ed. Jean-Louis Aroui and Andy Arleo. *Towards a Typology of Poetic Forms: From language to metrics and beyond.* Amsterdam: John Benjamins. 267-286.

Hayes, Bruce and Claire Moore-Cantwell. 2009. "Gerard Manley Hopkin's Sprung Rhythm: Corpus study and stochastic grammar." UCLA Phonology Seminar, October 13, 2009.

Kiparsky, Paul. 1977. "The Rhythmic Structure of English Verse." *Linguistic Inquiry*. 8(2): 189-248.

_____. 2005. "Iambic Inversion in Finnish." *A Man of Measure: Festschrift in Honour of Fred Karlsson*. 138-148.

Jesperson, Otto. 1933. "Notes on Metre." *Linguistica*. from ed. Seymour Chatman and Samuel R. Levin. 1967. *Essays on the Language of Literature*. Boston: Houghton Mifflin Company. 71-90.

Prince, Alan. 1989. "Metrical Forms." ed. Paul Kiparsky and Gilbert Youmans. *Phonetics and Phonology, Volume 1: Rhythm and Meter*. San Diego, CA: Academic Press, Inc. 45-80.

Shakespeare, William. 1974. *The Riverside Shakespeare*. 2nd ed. ed. G Blakemore Evans. Boston: Houghton Mifflin Company.